Approximating the covariance matrix with heavy tailed columns and RIP.

Alexander Litvak

University of Alberta

based on a joint work with

O. Guédon, A. Pajor and N. Tomczak-Jaegermann

(the paper "On the interval ..." available at: http://www.math.ualberta.ca/~alexandr/)

Aleksander Pełczyński Memorial Conference

Bedlewo, 2014

Notations

$\langle \cdot , \cdot \rangle$ denotes the canonical inner product on $\mathbb{R}^n$.

$| \cdot |$ denotes the canonical Euclidean norm on $\mathbb{R}^n$.

Notations

$\langle \cdot , \cdot \rangle$ denotes the canonical inner product on $\mathbb{R}^n$.
$| \cdot |$ denotes the canonical Euclidean norm on $\mathbb{R}^n$.

A random vector $X \in \mathbb{R}^n$ is called isotropic if for all $y \in \mathbb{R}^n$.

$$\mathbb{E}\langle X, y\rangle = 0 \quad \text{and} \quad \mathbb{E}\, |\langle X, y\rangle|^2 = |y|^2.$$

In other words, if X is centered and its covariance matrix is the identity:

$$\mathbb{E}\, X \otimes X = Id$$

(recall $\quad (X \otimes Y)(z) = \langle X, z\rangle Y \quad$ or $\quad X \otimes Y = \{Y_i X_j\}_{ij}$).

Notations

$\langle \cdot, \cdot \rangle$ denotes the canonical inner product on $\mathbb{R}^n$.
$| \cdot |$ denotes the canonical Euclidean norm on $\mathbb{R}^n$.

A random vector $X \in \mathbb{R}^n$ is called isotropic if for all $y \in \mathbb{R}^n$.

$$\mathbb{E}\langle X, y\rangle = 0 \quad \text{and} \quad \mathbb{E}\,|\langle X, y\rangle|^2 = |y|^2.$$

In other words, if X is centered and its covariance matrix is the identity:

$$\mathbb{E}\, X \otimes X = Id$$

(recall $\quad (X \otimes Y)(z) = \langle X, z\rangle Y \quad$ or $\quad X \otimes Y = \{Y_i X_j\}_{ij}$).

For an $n \times N$ matrix T its operator norm from ℓ_2^N to ℓ_2^n is denoted by

$$\|T\| = \sup_{|x|=1} |Tx|.$$

KLS problem

We consider the following model: $X_1, \ldots, X_N$ are independent random vectors in $\mathbb{R}^n$. For simplicity we assume that they are identically distributed and isotropic.

KLS problem

We consider the following model: $X_1, \dots, X_N$ are independent random vectors in $\mathbb{R}^n$. For simplicity we assume that they are identically distributed and isotropic.

Approximation of covariance matrix
(Kannan-Lovász-Simonovits (KLS) question):

How many random vectors X_i are needed for the empirical covariance matrix

$$\frac{1}{N}\sum_{i=1}^{N} X_i \otimes X_i$$

to approximate the identity with overwhelming probability?

(In Asymptotic Geometric Analysis this question was first asked about vectors uniformly distributed in an isotropic convex body. The approximation was needed in order to estimate the complexity of an algorithm computing the volume of the body).

KLS problem

Given $\varepsilon > 0$, how large N must be in order to have

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - \mathrm{Id} \right\|$$

KLS problem

Given $\varepsilon > 0$, how large N must be in order to have

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - \mathrm{Id} \right\| = \sup_{y \in S^{n-1}} \left| \frac{1}{N} \sum_{i=1}^{N} \left(\langle X_i, y \rangle^2 - 1 \right) \right|$$

KLS problem

Given $\varepsilon > 0$, how large N must be in order to have

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - \mathrm{Id} \right\| = \sup_{y \in S^{n-1}} \left| \frac{1}{N} \sum_{i=1}^{N} \left(\langle X_i, y \rangle^2 - 1 \right) \right| \leq \varepsilon$$

KLS problem

Given $\varepsilon > 0$, how large N must be in order to have

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - \mathrm{Id} \right\| = \sup_{y \in S^{n-1}} \left| \frac{1}{N} \sum_{i=1}^{N} \left(\langle X_i, y \rangle^2 - 1 \right) \right| \leq \varepsilon$$

or, equivalently,

$$\forall y \in S^{n-1} \qquad 1 - \varepsilon \leq \frac{1}{N} \sum_{i=1}^{N} \langle X_i, y \rangle^2 \leq 1 + \varepsilon.$$

KLS problem

Given $\varepsilon > 0$, how large N must be in order to have

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - \mathrm{Id} \right\| = \sup_{y \in S^{n-1}} \left| \frac{1}{N} \sum_{i=1}^{N} \left(\langle X_i, y \rangle^2 - 1 \right) \right| \leq \varepsilon$$

or, equivalently,

$$\forall y \in S^{n-1} \qquad 1 - \varepsilon \leq \frac{1}{N} \sum_{i=1}^{N} \langle X_i, y \rangle^2 \leq 1 + \varepsilon.$$

KLS (**95/97**): $N \sim C(\varepsilon, \delta) n^2$ with **Prob** $\geq 1 - \delta$.

KLS problem

Given $\varepsilon > 0$, how large N must be in order to have

$$\left\| \frac{1}{N}\sum_{i=1}^{N} X_i \otimes X_i - \mathrm{Id} \right\| = \sup_{y\in S^{n-1}} \left| \frac{1}{N}\sum_{i=1}^{N} \left(\langle X_i, y\rangle^2 - 1 \right) \right| \le \varepsilon$$

or, equivalently,

$$\forall y \in S^{n-1} \qquad 1-\varepsilon \le \frac{1}{N}\sum_{i=1}^{N} \langle X_i, y\rangle^2 \le 1+\varepsilon.$$

KLS (**95/97**): $N \sim C(\varepsilon,\delta) n^2$ with **Prob** $\ge 1-\delta$.

Bourgain (**96/99**): $N \sim C(\varepsilon,\delta) n \ln^3 n$ with **Prob** $\ge 1-\delta$.

KLS problem

Given $\varepsilon > 0$, how large N must be in order to have

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - \mathrm{Id} \right\| = \sup_{y \in S^{n-1}} \left| \frac{1}{N} \sum_{i=1}^{N} \left(\langle X_i, y \rangle^2 - 1 \right) \right| \le \varepsilon$$

or, equivalently,

$$\forall y \in S^{n-1} \qquad 1 - \varepsilon \le \frac{1}{N} \sum_{i=1}^{N} \langle X_i, y \rangle^2 \le 1 + \varepsilon.$$

KLS (**95/97**): $N \sim C(\varepsilon, \delta) n^2$ with **Prob** $\ge 1 - \delta$.

Bourgain (**96/99**): $N \sim C(\varepsilon, \delta) n \ln^3 n$ with **Prob** $\ge 1 - \delta$.

Improved to $N \sim C(\varepsilon, \delta) n \ln^2 n$ by Rudelson and to $N \sim C(\varepsilon, \delta) n \ln n$ by Giannopoulos, Hartzoulaki, Tsolomitis and by Paouris.

KLS problem

Given $\varepsilon > 0$, how large N must be in order to have

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - \mathrm{Id} \right\| = \sup_{y \in S^{n-1}} \left| \frac{1}{N} \sum_{i=1}^{N} \left(\langle X_i, y \rangle^2 - 1 \right) \right| \leq \varepsilon$$

or, equivalently,

$$\forall y \in S^{n-1} \qquad 1 - \varepsilon \leq \frac{1}{N} \sum_{i=1}^{N} \langle X_i, y \rangle^2 \leq 1 + \varepsilon.$$

KLS (**95/97**): $N \sim C(\varepsilon, \delta) n^2$ with **Prob** $\geq 1 - \delta$.

Bourgain (**96/99**): $N \sim C(\varepsilon, \delta) n \ln^3 n$ with **Prob** $\geq 1 - \delta$.

Improved to $N \sim C(\varepsilon, \delta) n \ln^2 n$ by Rudelson and to $N \sim C(\varepsilon, \delta) n \ln n$ by Giannopoulos, Hartzoulaki, Tsolomitis and by Paouris.

Aubrun (**07**): $N \sim n/\varepsilon^2$ if X_1 is unconditional with **Prob** $\geq 1 - \exp(-cn^{1/5})$.

Solution of KLS Problem in log-concave setting.

Theorem (Adamczak-LPT, 2010)

Let $X_1, \ldots, X_N$ be independent isotropic log-concave random vectors. Let $\varepsilon \in (0,1)$. Then for $N \geq Cn/\varepsilon^2$ *one has*

$$\mathbb{P}\left(\sup_{y \in S^{n-1}} \left| \frac{1}{N} \sum_{i=1}^{N} (|\langle X_i, y\rangle|^2 - \mathbb{E}|\langle X_i, y\rangle|^2) \right| \leq \varepsilon \right) \geq 1 - \exp\left(-c\sqrt{n}\right).$$

Solution of KLS Problem in log-concave setting.

Theorem (Adamczak-LPT, 2010)

Let $X_1, \ldots, X_N$ be independent isotropic log-concave random vectors. Let $\varepsilon \in (0,1)$. Then for $N \geq Cn/\varepsilon^2$ one has

$$\mathbb{P}\left(\sup_{y \in S^{n-1}} \left| \frac{1}{N} \sum_{i=1}^{N} (|\langle X_i, y\rangle|^2 - \mathbb{E}|\langle X_i, y\rangle|^2) \right| \leq \varepsilon \right) \geq 1 - \exp\left(-c\sqrt{n}\right).$$

In other words, for $N \geq Cn$, with high probability we have

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - Id \right\| \leq C\sqrt{\frac{n}{N}}.$$

Solution of KLS Problem in log-concave setting.

Theorem (Adamczak-LPT, 2010)

Let $X_1, \ldots, X_N$ be independent isotropic log-concave random vectors. Let $\varepsilon \in (0,1)$. Then for $N \geq Cn/\varepsilon^2$ *one has*

$$\mathbb{P}\left(\sup_{y \in S^{n-1}} \left| \frac{1}{N} \sum_{i=1}^{N} (|\langle X_i, y\rangle|^2 - \mathbb{E}|\langle X_i, y\rangle|^2) \right| \leq \varepsilon \right) \geq 1 - \exp\left(-c\sqrt{n}\right).$$

In other words, for $N \geq Cn$, with high probability we have

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - Id \right\| \leq C\sqrt{\frac{n}{N}}.$$

Remark. A measure μ on $\mathbb{R}^n$ is log-concave if for every measurable $A, B \subset \mathbb{R}^n$ and every $\theta \in [0,1]$,

$$\mu(\theta A + (1-\theta)B) \geq \mu(A)^{\theta} \mu(B)^{(1-\theta)}$$

Relations to standard Random Matrix Theory (RMT)

RMT studies in particular limit behavior of singular numbers of random matrices. Recall for $n \times N$ matrix A, the largest and the smallest singular values are defined as

$$s_1(A) = \sup_{|x|=1} \|Ax\| = \|A\| \quad \text{and} \quad s_n(A) = \inf_{|x|=1} \|Ax\| = 1/\|A^{-1}\|.$$

Classical result is the Bai-Yin Theorem.

Theorem (Bai-Yin)

Let A be an $n \times N$ random matrix with i.i.d. entries whose 4-th moments are bounded. Let

$$\beta = \lim_{n\to\infty} \frac{n}{N} \in (0,1).$$

Then

$$1 - \sqrt{\beta} = \lim_{n\to\infty} s_n(A/\sqrt{N}) \le \lim_{n\to\infty} s_1(A/\sqrt{N}) = 1 + \sqrt{\beta}.$$

Relations to standard Random Matrix Theory (RMT)

RMT studies in particular limit behavior of singular numbers of random matrices. Recall for $n \times N$ matrix A, the largest and the smallest singular values are defined as

$$s_1(A) = \sup_{|x|=1} \|Ax\| = \|A\| \quad \text{and} \quad s_n(A) = \inf_{|x|=1} \|Ax\| = 1/\|A^{-1}\|.$$

Classical result is the Bai-Yin Theorem.

Theorem (Bai-Yin)

Let A be an $n \times N$ random matrix with i.i.d. entries whose 4-th moments are bounded. Let

$$\beta = \lim_{n\to\infty} \frac{n}{N} \in (0,1).$$

Then

$$1 - \sqrt{\beta} = \lim_{n\to\infty} s_n(A/\sqrt{N}) \leq \lim_{n\to\infty} s_1(A/\sqrt{N}) = 1 + \sqrt{\beta}.$$

AGA point of view: We are interested in **asymptotic non-limit** behavior, i.e. we would like to provide the quantitative estimates on the rate of convergence.

Relations to standard RMT

Theorem (ALPT)

Let $n \leq N$. Let A be a random $n \times N$ matrix, whose columns $X_1, \ldots, X_N$ are isotropic log-concave independent random vectors in $\mathbb{R}^n$. Denoting $\beta = n/N$ we have

$$1 - C\sqrt{\beta} \leq s_n(A/\sqrt{N}) \leq s_1(A/\sqrt{N}) \leq 1 + C\sqrt{\beta},$$

with probability at least $1 - 2\exp(-c\sqrt{n})$.

Relations to standard RMT

Theorem (ALPT)

Let $n \le N$. Let A be a random $n \times N$ matrix, whose columns $X_1, \dots, X_N$ are isotropic log-concave independent random vectors in $\mathbb{R}^n$. Denoting $\beta = n/N$ we have

$$1 - C\sqrt{\beta} \le s_n(A/\sqrt{N}) \le s_1(A/\sqrt{N}) \le 1 + C\sqrt{\beta},$$

with probability at least $1 - 2\exp(-c\sqrt{n})$.

Compare with the Bai-Yin Theorem:

$$1 - \sqrt{\beta} = \lim_{n\to\infty} s_n \le \lim_{n\to\infty} s_1 = 1 + \sqrt{\beta}.$$

Approximation of covariance matrix: heavy tails

Question: Under what conditions can the KLS problem be solved with $N \sim n$?

Approximation of covariance matrix: heavy tails

Question: Under what conditions can the KLS problem be solved with $N \sim n$?

For example, is it enough to assume that $|X_i| \le C\sqrt{n}$ with high probability and

$$\sigma_q(X_1) := \sup_{|x|=1} \left(\mathbb{E}|\langle X_1, x\rangle|^q\right)^{1/q} \le C$$

for some $q > 2$?

Approximation of covariance matrix: heavy tails

Question: Under what conditions can the KLS problem be solved with $N \sim n$?

For example, is it enough to assume that $|X_i| \leq C\sqrt{n}$ with high probability and

$$\sigma_q(X_1) := \sup_{|x|=1} \left(\mathbb{E}|\langle X_1, x\rangle|^q\right)^{1/q} \leq C$$

for some $q > 2$?

Vershynin (**2012**): For $q > 4$ if $\sigma_q(X_1) \leq C_1$ and if $|X_i| < C_1\sqrt{n}$ a.s. then

$$\left\|\frac{1}{N}\sum_{i=1}^{N} X_i \otimes X_i - \mathrm{Id}\right\| \leq C\,(\ln\ln n)^2\,(n/N)^{1/2-1/q}.$$

Approximation of covariance matrix: heavy tails

Question: Under what conditions can the KLS problem be solved with $N \sim n$?

For example, is it enough to assume that $|X_i| \leq C\sqrt{n}$ with high probability and

$$\sigma_q(X_1) := \sup_{|x|=1} \left(\mathbb{E}|\langle X_1, x\rangle|^q\right)^{1/q} \leq C$$

for some $q > 2$?

Vershynin (**2012**): For $q > 4$ if $\sigma_q(X_1) \leq C_1$ and if $|X_i| < C_1\sqrt{n}$ a.s. then

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - \mathrm{Id} \right\| \leq C\,(\ln\ln n)^2\,(n/N)^{1/2-1/q}.$$

He also conjectured that

"$\ln\ln n$ is not needed for an appropriate q, probably $q = 4$ or even $q > 2$."

Approximation of covariance matrix: heavy tails

Srivastava and Vershynin (**2013**): A solution (in average) under strong assumption on projections: there is $\eta > 0$ such that for every projection of rank k and every $t \geq C\sqrt{k}$,

$$\mathbb{P}\left(|PX_1| \geq t\right) \leq C/t^{2(1+\eta)}$$

Approximation of covariance matrix: heavy tails

Srivastava and Vershynin (**2013**): A solution (in average) under strong assumption on projections: there is $\eta > 0$ such that for every projection of rank k and every $t \geq C\sqrt{k}$,

$$\mathbb{P}\left(|PX_1| \geq t\right) \leq C/t^{2(1+\eta)}$$

Moreover, for the smallest singular value only 1-dimensional projections are needed: $\sigma_q(X_1) \leq C$ for some $q > 2$.

Approximation of covariance matrix: heavy tails

Srivastava and Vershynin (**2013**): A solution (in average) under strong assumption on projections: there is $\eta > 0$ such that for every projection of rank k and every $t \geq C\sqrt{k}$,

$$\mathbb{P}\left(|PX_1| \geq t\right) \leq C/t^{2(1+\eta)}$$

Moreover, for the smallest singular value only 1-dimensional projections are needed: $\sigma_q(X_1) \leq C$ for some $q > 2$.

Mendelson and Paouris (**2012, 2014**): A solution with high probability

1. For $q > 4$ assuming that X_1 is unconditional and that for some $p > 2$

$$\exists\, p > 2: \quad \|X_1\|_{\ell_p^n} \leq C n^{1/p} \ \text{ a.s.}$$

Approximation of covariance matrix: heavy tails

Srivastava and Vershynin (**2013**): A solution (in average) under strong assumption on projections: there is $\eta > 0$ such that for every projection of rank k and every $t \geq C\sqrt{k}$,

$$\mathbb{P}\left(|PX_1| \geq t\right) \leq C/t^{2(1+\eta)}$$

Moreover, for the smallest singular value only 1-dimensional projections are needed: $\sigma_q(X_1) \leq C$ for some $q > 2$.

Mendelson and Paouris (**2012, 2014**): A solution with high probability

1. For $q > 4$ assuming that X_1 is unconditional and that for some $p > 2$

$$\exists\, p > 2: \quad \|X_1\|_{\ell_p^n} \leq C n^{1/p} \ \text{ a.s.}$$

2. For $q > 8$ assuming that $\sigma_q(X_1) \leq C$ and that $\max_i |X_i| \leq (nN)^{1/4}$ a.s.

Approximation of covariance matrix: heavy tails

Srivastava and Vershynin (**2013**): A solution (in average) under strong assumption on projections: there is $\eta > 0$ such that for every projection of rank k and every $t \geq C\sqrt{k}$,

$$\mathbb{P}\left(|PX_1| \geq t\right) \leq C/t^{2(1+\eta)}$$

Moreover, for the smallest singular value only 1-dimensional projections are needed: $\sigma_q(X_1) \leq C$ for some $q > 2$.

Mendelson and Paouris (**2012, 2014**): A solution with high probability

1. For $q > 4$ assuming that X_1 is unconditional and that for some $p > 2$

$$\exists\, p > 2 : \quad \|X_1\|_{\ell_p^n} \leq C n^{1/p} \ \text{ a.s.}$$

2. For $q > 8$ assuming that $\sigma_q(X_1) \leq C$ and that $\quad \max_i |X_i| \leq (nN)^{1/4}$ a.s.

In both MP and SV works: for i.i.d. entries with bounded moment $q > 4$.

KLS: 4-th moment

Theorem (GLPT)

Let $X_1, \dots, X_N$ be independent isotropic random vectors. Let $4 < q \leq 8$ and $p < q - 4$. Assume that

$$\forall y \in S^{n-1} \quad \forall t > 0 \qquad \mathbb{P}\left(|\langle X, y\rangle| > t\right) \leq t^{-q}.$$

Then with high probability

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - Id \right\| \leq \frac{C}{N} \max_{i \leq N} |X_i|^2 + C(p,q) \left(\frac{n}{N}\right)^{p/q}.$$

KLS: 4-th moment

Theorem (GLPT)

Let $X_1, \dots, X_N$ be independent isotropic random vectors. Let $4 < q \le 8$ and $p < q - 4$. Assume that

$$\forall y \in S^{n-1} \quad \forall t > 0 \qquad \mathbb{P}\left(|\langle X, y\rangle| > t\right) \le t^{-q}.$$

Then with high probability

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - Id \right\| \le \frac{C}{N} \max_{i \le N} |X_i|^2 + C(p,q) \left(\frac{n}{N}\right)^{p/q}.$$

In particular,

if $\quad \max_{i \le N} |X_i|^2 \le n^{p/q} N^{1-p/q} \quad$ *then* $\quad \left\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - Id \right\| \le C(p,q) \left(\frac{n}{N}\right)^{p/q}.$

RIP

Candes and Tao (**2005**) introduced the concept of Restricted Isometry Property (**RIP**) for a given matrix T in search for sufficient conditions for T to satisfy some "reconstruction" conditions from "compressed sensing" and coding theory.

RIP

Candes and Tao (**2005**) introduced the concept of Restricted Isometry Property (**RIP**) for a given matrix T in search for sufficient conditions for T to satisfy some "reconstruction" conditions from "compressed sensing" and coding theory.

RIP parameter of order m

is the smallest number $\delta = \delta_m(T)$ such that for every m-sparse vector $x \in \mathbb{R}^N$

$$(1-\delta)|x|^2 \leq |Tx|^2 \leq (1+\delta)|x|^2$$

(x is m-sparse if it has at most m non-zero coordinates).

RIP

Candes and Tao (**2005**) introduced the concept of Restricted Isometry Property (**RIP**) for a given matrix T in search for sufficient conditions for T to satisfy some "reconstruction" conditions from "compressed sensing" and coding theory.

RIP parameter of order m

is the smallest number $\delta = \delta_m(T)$ such that for every m-sparse vector $x \in \mathbb{R}^N$

$$(1-\delta)|x|^2 \leq |Tx|^2 \leq (1+\delta)|x|^2$$

(x is m-sparse if it has at most m non-zero coordinates).

That is, every sub-matrix of T obtained by taking m columns is "almost" isometry.

RIP

Candes and Tao (**2005**) introduced the concept of Restricted Isometry Property (**RIP**) for a given matrix T in search for sufficient conditions for T to satisfy some "reconstruction" conditions from "compressed sensing" and coding theory.

RIP parameter of order m

is the smallest number $\delta = \delta_m(T)$ such that for every m-sparse vector $x \in \mathbb{R}^N$

$$(1-\delta)|x|^2 \leq |Tx|^2 \leq (1+\delta)|x|^2$$

(x is m-sparse if it has at most m non-zero coordinates).

That is, every sub-matrix of T obtained by taking m columns is "almost" isometry.

Note that columns of T have to satisfy $|T_i| = |Te_i| \sim 1$.

RIP

Candes and Tao (**2005**) introduced the concept of Restricted Isometry Property (**RIP**) for a given matrix T in search for sufficient conditions for T to satisfy some "reconstruction" conditions from "compressed sensing" and coding theory.

RIP parameter of order m

is the smallest number $\delta = \delta_m(T)$ such that for every m-sparse vector $x \in \mathbb{R}^N$

$$(1-\delta)|x|^2 \leq |Tx|^2 \leq (1+\delta)|x|^2$$

(x is m-sparse if it has at most m non-zero coordinates).

That is, every sub-matrix of T obtained by taking m columns is "almost" isometry.

Note that columns of T have to satisfy $|T_i| = |Te_i| \sim 1$.

There are many papers on these topics.

RIP

Candes proved that if a matrix T satisfies

$$\delta_{2m}(T) < \delta_0 = \sqrt{2} - 1$$

then the following holds:

RIP

Candes proved that if a matrix T satisfies

$$\delta_{2m}(T) < \delta_0 = \sqrt{2} - 1$$

then the following holds:

basis pursuit algorithm (exact reconstruction by ℓ_1 minimization)

whenever $Tz = y$ has an m-sparse solution z_0, then z_0 is the unique solution of the problem

$$\min \|z\|_1, \quad Tz = y.$$

RIP

Candes proved that if a matrix T satisfies

$$\delta_{2m}(T) < \delta_0 = \sqrt{2} - 1$$

then the following holds:

basis pursuit algorithm (exact reconstruction by ℓ_1 minimization)

whenever $Tz = y$ has an m-sparse solution z_0, then z_0 is the unique solution of the problem

$$\min \|z\|_1, \quad Tz = y.$$

Donoho (2005) showed that the later condition is equivalent to a condition on the neighborliness of polytopes in $\mathbb{R}^n$ (i.e. the above is equivalent to the following: TB_1^N is m-centrally-neighborly, that is every set of m vertices containing no opposite pairs forms a vertex set of a face).

RIP

Let $X_1, ..., X_N$ be i.i.d. random vectors in $\mathbb{R}^n$ and assume that their Euclidean norms are concentrated around $\sqrt{n}$. Let T be an $n \times N$ matrix whose columns are $X_i/\sqrt{n}$.

RIP

Let $X_1, ..., X_N$ be i.i.d. random vectors in $\mathbb{R}^n$ and assume that their Euclidean norms are concentrated around $\sqrt{n}$. Let T be an $n \times N$ matrix whose columns are $X_i/\sqrt{n}$.

In the Gaussian case, the Bernoulli (± 1) case, the sub-Gaussian case one has $\delta_{2m} \leq \delta_0$ with high probability for

$$m = \frac{Cn}{\ln(2N/n)}.$$

Many works by: Baraniuk, Candes, Cohen, Dahmen, Davenport, DeVore, Donoho, Kashin, Mendelson, Pajor, Romberg, Rudelson, Tao, Temlyakov, Vershynin, Tomczak-Jaegermann, Wakin...

RIP

Let $X_1, ..., X_N$ be i.i.d. random vectors in $\mathbb{R}^n$ and assume that their Euclidean norms are concentrated around $\sqrt{n}$. Let T be an $n \times N$ matrix whose columns are $X_i/\sqrt{n}$.

In the Gaussian case, the Bernoulli (± 1) case, the sub-Gaussian case one has $\delta_{2m} \le \delta_0$ with high probability for

$$m = \frac{Cn}{\ln(2N/n)}.$$

Many works by: Baraniuk, Candes, Cohen, Dahmen, Davenport, DeVore, Donoho, Kashin, Mendelson, Pajor, Romberg, Rudelson, Tao, Temlyakov, Vershynin, Tomczak-Jaegermann, Wakin...

ALPT **(2011)**: Similar estimates with $\quad m = Cn/\ln^2(2N/n) \quad$ provided that

$$\forall y \in S^{n-1} \qquad \mathbb{P}\left(|\langle X_1, y\rangle| > t\right) \le C\exp(-ct).$$

For example, for isotropic log-concave vectors.

RIP

Let $X_1, \dots, X_N$ be i.i.d. random vectors in $\mathbb{R}^n$ and assume that their Euclidean norms are concentrated around $\sqrt{n}$. Let T be an $n \times N$ matrix whose columns are $X_i/\sqrt{n}$.

In the Gaussian case, the Bernoulli (± 1) case, the sub-Gaussian case one has $\delta_{2m} \le \delta_0$ with high probability for

$$m = \frac{Cn}{\ln(2N/n)}.$$

Many works by: Baraniuk, Candes, Cohen, Dahmen, Davenport, DeVore, Donoho, Kashin, Mendelson, Pajor, Romberg, Rudelson, Tao, Temlyakov, Vershynin, Tomczak-Jaegermann, Wakin...

ALPT **(2011)**: Similar estimates with $\quad m = Cn/\ln^2(2N/n) \quad$ provided that

$$\forall y \in S^{n-1} \qquad \mathbb{P}\left(|\langle X_1, y\rangle| > t\right) \le C\exp(-ct).$$

For example, for isotropic log-concave vectors.

Question. Under what (weakest) conditions on X_i's can one obtain RIP?

Main Results

Theorem (GLPT)

Let $q > 4$ and $p > \frac{4}{q-4}$. Let $X_1, ..., X_N$ be independent random vectors in $\mathbb{R}^n$ such that that their Euclidean norms are concentrated around $\sqrt{n}$ and assume

$$\forall y \in S^{n-1} \quad \mathbb{P}\left(|\langle X_i, y\rangle| > t\right) \le C/t^q.$$

Then the matrix T whose columns are $X_i/\sqrt{n}$ satisfies $\delta_{2m}(T) \le \delta_0$ with high probability for

$$m = C(p,q)\,\frac{n}{(N/n)^p}.$$

Main Results

Theorem (GLPT)

Let $q > 4$ and $p > \frac{4}{q-4}$. Let $X_1, ..., X_N$ be independent random vectors in $\mathbb{R}^n$ such that that their Euclidean norms are concentrated around $\sqrt{n}$ and assume

$$\forall y \in S^{n-1} \quad \mathbb{P}\left(|\langle X_i, y\rangle| > t\right) \leq C/t^q.$$

Then the matrix T whose columns are $X_i/\sqrt{n}$ satisfies $\delta_{2m}(T) \leq \delta_0$ with high probability for

$$m = C(p,q)\,\frac{n}{(N/n)^p}.$$

Sharpness: For $p < \frac{2}{q-2}$, one can't get better than

$$C_0(p,q)\,\frac{n}{(N/n)^p}.$$

Main Results

Theorem (GLPT)

Let $\alpha \in (0,2]$. Let $X_1, ..., X_N$ be independent random vectors in $\mathbb{R}^n$ such that that their Euclidean norms are concentrated around $\sqrt{n}$ and assume

$$\forall y \in S^{n-1} \qquad \mathbb{P}\left(|\langle X_i, y\rangle| > t\right) \le C\exp(-ct^{\alpha}).$$

Then the matrix T whose columns are $X_i/\sqrt{n}$ satisfies $\delta_{2m}(T) \le \delta_0$ with high probability for

$$m = \frac{C_\alpha\, n}{(\ln(2N/n))^{2/\alpha}}.$$

Sharpness: The bound on m is sharp up to constant C_α.

Ideas of proofs

The main technical tool is obtaining bounds on the following two parameters.

Ideas of proofs

The main technical tool is obtaining bounds on the following two parameters.

For $m \leq N$ and random vectors X_1, ..., X_N in $\mathbb{R}^n$, define A_m and B_m by

1.

$$A_m := \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \leq m}} \left| \sum_{i=1}^{N} a_i X_i \right| .$$

Ideas of proofs

The main technical tool is obtaining bounds on the following two parameters.

For $m \le N$ and random vectors X_1, ..., X_N in $\mathbb{R}^n$, define A_m and B_m by

1.

$$A_m := \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \le m}} \left| \sum_{i=1}^{N} a_i X_i \right| .$$

Note, if A is the matrix with columns X_i, then A_m is the supremum of norms of submatrices consisting of m columns of A. The problem of estimating A_m is interesting by itself, although for KLS problem only $m = n$ is needed.

Ideas of proofs

The main technical tool is obtaining bounds on the following two parameters.

For $m \leq N$ and random vectors X_1, ..., X_N in $\mathbb{R}^n$, define A_m and B_m by

1.

$$A_m := \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \leq m}} \left| \sum_{i=1}^{N} a_i X_i \right| .$$

Note, if A is the matrix with columns X_i, then A_m is the supremum of norms of submatrices consisting of m columns of A. The problem of estimating A_m is interesting by itself, although for KLS problem only $m = n$ is needed.

2.

$$B_m^2 := \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \leq m}} \left| \left| \sum_{i=1}^{N} a_i X_i \right|^2 - \sum_{i=1}^{N} a_i^2 |X_i|^2 \right|$$

Ideas of proofs

The main technical tool is obtaining bounds on the following two parameters.

For $m \leq N$ and random vectors X_1, ..., X_N in $\mathbb{R}^n$, define A_m and B_m by

1.

$$A_m := \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \leq m}} \left| \sum_{i=1}^{N} a_i X_i \right| .$$

Note, if A is the matrix with columns X_i, then A_m is the supremum of norms of submatrices consisting of m columns of A. The problem of estimating A_m is interesting by itself, although for KLS problem only $m = n$ is needed.

2.

$$B_m^2 := \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \leq m}} \left| \left| \sum_{i=1}^{N} a_i X_i \right|^2 - \sum_{i=1}^{N} a_i^2 |X_i|^2 \right| = \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \leq m}} \left| \sum_{i \neq j} \langle a_i X_i, a_j X_j \rangle \right| .$$

Ideas of proofs

The main technical tool is obtaining bounds on the following two parameters.

For $m \leq N$ and random vectors X_1, ..., X_N in $\mathbb{R}^n$, define A_m and B_m by

1.

$$A_m := \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \leq m}} \left| \sum_{i=1}^{N} a_i X_i \right| .$$

Note, if A is the matrix with columns X_i, then A_m is the supremum of norms of submatrices consisting of m columns of A. The problem of estimating A_m is interesting by itself, although for KLS problem only $m = n$ is needed.

2.

$$B_m^2 := \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \leq m}} \left| \left| \sum_{i=1}^{N} a_i X_i \right|^2 - \sum_{i=1}^{N} a_i^2 |X_i|^2 \right| = \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \leq m}} \left| \sum_{i \neq j} \langle a_i X_i, a_j X_j \rangle \right| .$$

B_m is related to concentration. An upper bound on it plays the crucial role for RIP.

RIP

RIP parameter δ_m can be rewritten as

$$\delta_m(A/\sqrt{n}) = \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \leq m}} \left| \frac{1}{n} |Aa|^2 - 1 \right|$$

RIP

RIP parameter δ_m can be rewritten as

$$\delta_m(A/\sqrt{n}) = \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \le m}} \left| \frac{1}{n} |Aa|^2 - 1 \right|$$

$$\le \frac{1}{n} \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \le m}} \left| \frac{1}{n} |Aa|^2 - \frac{1}{n} \sum_{i=1}^{N} a_i^2 |X_i|^2 \right| + \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \le m}} \left| \frac{1}{n} \sum_{i=1}^{N} a_i^2 |X_i|^2 - 1 \right|$$

RIP

RIP parameter δ_m can be rewritten as

$$\delta_m(A/\sqrt{n}) = \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \leq m}} \left| \frac{1}{n} |Aa|^2 - 1 \right|$$

$$\leq \frac{1}{n} \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \leq m}} \left| \frac{1}{n} |Aa|^2 - \frac{1}{n} \sum_{i=1}^{N} a_i^2 |X_i|^2 \right| + \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \leq m}} \left| \frac{1}{n} \sum_{i=1}^{N} a_i^2 |X_i|^2 - 1 \right|$$

$$\leq \frac{1}{n} B_m^2 + \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \leq m}} \left| \sum_{i=1}^{N} a_i^2 \left(\frac{1}{n} |X_i|^2 - 1 \right) \right|$$

RIP

RIP parameter δ_m can be rewritten as

$$\delta_m(A/\sqrt{n}) = \sup_{\substack{a\in S^{N-1} \\ |\operatorname{supp}(a)|\le m}} \left| \frac{1}{n}|Aa|^2 - 1 \right|$$

$$\le \frac{1}{n} \sup_{\substack{a\in S^{N-1} \\ |\operatorname{supp}(a)|\le m}} \left| \frac{1}{n}|Aa|^2 - \frac{1}{n}\sum_{i=1}^N a_i^2 |X_i|^2 \right| + \sup_{\substack{a\in S^{N-1} \\ |\operatorname{supp}(a)|\le m}} \left| \frac{1}{n}\sum_{i=1}^N a_i^2 |X_i|^2 - 1 \right|$$

$$\le \frac{1}{n} B_m^2 + \sup_{\substack{a\in S^{N-1} \\ |\operatorname{supp}(a)|\le m}} \left| \sum_{i=1}^N a_i^2 \left(\frac{1}{n}|X_i|^2 - 1 \right) \right| \le B_m^2 + \max_{i\le N} \left| \frac{1}{n}|X_i|^2 - 1 \right|.$$

RIP

RIP parameter δ_m can be rewritten as

$$\delta_m(A/\sqrt{n}) = \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \le m}} \left| \frac{1}{n} |Aa|^2 - 1 \right|$$

$$\le \frac{1}{n} \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \le m}} \left| \frac{1}{n} |Aa|^2 - \frac{1}{n} \sum_{i=1}^{N} a_i^2 |X_i|^2 \right| + \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \le m}} \left| \frac{1}{n} \sum_{i=1}^{N} a_i^2 |X_i|^2 - 1 \right|$$

$$\le \frac{1}{n} B_m^2 + \sup_{\substack{a \in S^{N-1} \\ |\operatorname{supp}(a)| \le m}} \left| \sum_{i=1}^{N} a_i^2 \left(\frac{1}{n} |X_i|^2 - 1 \right) \right| \le B_m^2 + \max_{i \le N} \left| \frac{1}{n} |X_i|^2 - 1 \right|.$$

Note that,

$$\max_{i \le N} \left| \frac{1}{n} |X_i|^2 - 1 \right| = \delta_1(A/\sqrt{n}) \le \delta_m(A/\sqrt{n}),$$

that is, concentration of $|X_i|$ around $\sqrt{n}$ is needed.

KLS

We need to estimate

$$\mathbb{P}\left(\sup_{a\in S^{n-1}}\left|\sum_{i=1}^{N}\left(\langle X_i,a\rangle^2-\mathbb{E}\langle X_i,a\rangle^2\right)\right|>t\right).$$

KLS

We need to estimate

$$\mathbb{P}\left(\sup_{a\in S^{n-1}}\left|\sum_{i=1}^{N}\left(\langle X_i,a\rangle^2-\mathbb{E}\langle X_i,a\rangle^2\right)\right|>t\right).$$

First, using symmetrization we pass to

$$\mathbb{P}\left(\sup_{a\in S^{n-1}}\left|\sum_{i=1}^{N}\varepsilon_i\langle X_i,a\rangle^2\right|>t\right),$$

where ε_i are independent Bernoulli ± 1 random variables.

KLS

We need to estimate

$$\mathbb{P}\left(\sup_{a\in S^{n-1}}\left|\sum_{i=1}^{N}\left(\langle X_i,a\rangle^2-\mathbb{E}\langle X_i,a\rangle^2\right)\right|>t\right).$$

First, using symmetrization we pass to

$$\mathbb{P}\left(\sup_{a\in S^{n-1}}\left|\sum_{i=1}^{N}\varepsilon_i\langle X_i,a\rangle^2\right|>t\right),$$

where ε_i are independent Bernoulli ± 1 random variables.

Conditioning on X_i and considering decreasing rearrangement,

$$\left|\sum_{i=1}^{N}\varepsilon_i\langle X_i,a\rangle^2\right|\le\sum_{i=1}^{m}\langle X_i,a\rangle^{*\,2}+\left|\sum_{i=m+1}^{N}\varepsilon_{\pi(i)}\langle X_i,a\rangle^{*\,2}\right|,$$

for some permutation π.

KLS

Now,

$$\sum_{i=1}^{m} \langle X_i, a \rangle^{*\,2} \le A_m^2$$

and using Hoeffding's inequality, for every $t > 0$

$$\mathbb{P}_{(\varepsilon_i)} \left(\left| \sum_{i=m+1}^{N} \varepsilon_{\pi(i)} \langle X_i, a \rangle^{*\,2} \right| \ge t \sqrt{\sum_{i=m+1}^{N} \langle X_i, a \rangle^{*\,4}} \right) \le 2 \exp(-t^2/2).$$

KLS

Now,

$$\sum_{i=1}^{m} \langle X_i, a\rangle^{*\,2} \le A_m^2$$

and using Hoeffding's inequality, for every $t > 0$

$$\mathbb{P}_{(\varepsilon_i)} \left(\left| \sum_{i=m+1}^{N} \varepsilon_{\pi(i)} \langle X_i, a\rangle^{*\,2} \right| \ge t \sqrt{\sum_{i=m+1}^{N} \langle X_i, a\rangle^{*\,4}} \right) \le 2\exp(-t^2/2).$$

Finally we estimate

$$\mathbb{P} \left(\sum_{i=m+1}^{N} \langle X_i, a\rangle^{*\,4} > s \right)$$

and choose parameters appropriately ($m = n$, $t = \sqrt{n}$, ...).

Bounds on A_m, B_m

Using decoupling argument,

$$\left|\sum_{i\neq j}\langle a_iX_i, a_jX_j\rangle\right| = 2^{2-N}\left|\sum_{I\subset\{1,2,\ldots,N\}}\langle\sum_{i\in I} a_iX_i, \sum_{j\in I^c} a_jX_j\rangle\right|.$$

Bounds on A_m, B_m

Using decoupling argument,

$$\left|\sum_{i\neq j}\langle a_iX_i, a_jX_j\rangle\right| = 2^{2-N}\left|\sum_{I\subset\{1,2,\ldots,N\}}\langle\sum_{i\in I}a_iX_i, \sum_{j\in I^c}a_jX_j\rangle\right|.$$

We denote

$$Q(a,I,I^c) := \left|\langle\sum_{i\in I}a_iX_i, \sum_{j\in I^c}a_jX_j\rangle\right| \qquad \text{and} \qquad Q_m(I) = \sup_{\substack{a\in S^{N-1} \\ |\operatorname{supp}(a)|\leq m}} Q(a,I,I^c).$$

Bounds on A_m, B_m

Using decoupling argument,

$$\left|\sum_{i\neq j}\langle a_iX_i, a_jX_j\rangle\right| = 2^{2-N}\left|\sum_{I\subset\{1,2,\ldots,N\}}\langle\sum_{i\in I}a_iX_i, \sum_{j\in I^c}a_jX_j\rangle\right|.$$

We denote

$$Q(a,I,I^c) := \left|\langle\sum_{i\in I}a_iX_i, \sum_{j\in I^c}a_jX_j\rangle\right| \qquad \text{and} \qquad Q_m(I) = \sup_{\substack{a\in S^{N-1}\\ |\operatorname{supp}(a)|\leq m}} Q(a,I,I^c).$$

Therefore

$$B_m^2 = \sup_{\substack{a\in S^{N-1}\\ |\operatorname{supp}(a)|\leq m}}\left|\sum_{i\neq j}\langle a_iX_i, a_jX_j\rangle\right| \leq 2^{2-N}\sum_{I\subset\{1,2,\ldots,N\}}Q_m(I).$$

Bounds on A_m, B_m

We prove that for some $\gamma \in [1/2, 1)$ and every $\varepsilon \in (0,1)$, $t > 1$, with high probability

$$Q_m(I) \leq (1+\varepsilon)(Q_{\gamma m}(I) + tA_m).$$

Bounds on A_m, B_m

We prove that for some $\gamma \in [1/2, 1)$ and every $\varepsilon \in (0, 1)$, $t > 1$, with high probability

$$Q_m(I) \leq (1+\varepsilon)(Q_{\gamma m}(I) + tA_m).$$

The we use iteration procedure choosing appropriate ε and t on every step and controlling probability. It will give a bound of the type

$$Q_m(I) \leq \prod_{\ell=1}^{k}(1+\varepsilon_\ell)\left(Q_{\gamma^k m}(I) + A_m \sum_{\ell=1}^{k} t_\ell\right)$$

Bounds on A_m, B_m

We prove that for some $\gamma \in [1/2, 1)$ and every $\varepsilon \in (0,1)$, $t > 1$, with high probability

$$Q_m(I) \le (1+\varepsilon)(Q_{\gamma m}(I) + tA_m).$$

The we use iteration procedure choosing appropriate ε and t on every step and controlling probability. It will give a bound of the type

$$Q_m(I) \le \prod_{\ell=1}^{k}(1+\varepsilon_\ell)\left(Q_{\gamma^k m}(I) + A_m \sum_{\ell=1}^{k} t_\ell\right) \le C(\max |X_i|^2 + \sqrt{m}\,(N/m)^{\beta} A_m),$$

Bounds on A_m, B_m

We prove that for some $\gamma \in [1/2, 1)$ and every $\varepsilon \in (0, 1)$, $t > 1$, with high probability

$$Q_m(I) \leq (1+\varepsilon)(Q_{\gamma m}(I) + tA_m).$$

The we use iteration procedure choosing appropriate ε and t on every step and controlling probability. It will give a bound of the type

$$Q_m(I) \leq \prod_{\ell=1}^{k}(1+\varepsilon_\ell)\left(Q_{\gamma^k m}(I) + A_m \sum_{\ell=1}^{k} t_\ell\right) \leq C(\max|X_i|^2 + \sqrt{m}\,(N/m)^\beta A_m),$$

which leads to

$$B_m^2 \leq C_1 \left(\max|X_i|^2 + \sqrt{m}(N/m)^\beta A_m\right)$$

Bounds on A_m, B_m

We prove that for some $\gamma \in [1/2, 1)$ and every $\varepsilon \in (0, 1)$, $t > 1$, with high probability

$$Q_m(I) \leq (1+\varepsilon)(Q_{\gamma m}(I) + tA_m).$$

The we use iteration procedure choosing appropriate ε and t on every step and controlling probability. It will give a bound of the type

$$Q_m(I) \leq \prod_{\ell=1}^{k}(1+\varepsilon_\ell)\left(Q_{\gamma^k m}(I) + A_m \sum_{\ell=1}^{k} t_\ell\right) \leq C(\max|X_i|^2 + \sqrt{m}\,(N/m)^\beta A_m),$$

which leads to

$$B_m^2 \leq C_1 \left(\max|X_i|^2 + \sqrt{m}(N/m)^\beta A_m\right)$$

$$\leq C_2 \left(\max|X_i|^2 + \sqrt{m}(N/m)^\beta B_m + \sqrt{m}(N/m)^\beta \ \max|X_i|\right).$$